A Study of Enhanced Robot Autonomy in Telepresence

A Study of Enhanced Robot Autonomy in Telepresence

Lorenzo Riano, Chris Burbridge, and T. M. McGinnity

Intelligent Systems Research Centre, University of Ulster, Londonderry, BT48 7JL, UK
{cj.burbridge, l.riano, tm.mcginnity}@ulster.ac.uk

Abstract. Many telepresence robots have been developed, but most offer little more than a webcam on wheels, with the user being in control of every function of the robot. In order to enhance the users' ability to remotely communicate, navigate and *feel* present in a remote environment, we argue that a higher degree of autonomy is required. In this paper a telepresence system is presented that assists the user in three core ways: semi-autonomous navigation control, semi-autonomous people tracking and improved situational awareness. The value of enhanced autonomy in telepresence systems is demonstrated by experiments carried out with real users, both with a webcam on wheels and with our proposed telepresence system.

1 Introduction

Telepresence robots give users a virtual presence at a remote location, allowing them to carry out tasks remotely. The nature of the task varies according to the specific application, ranging from homecare tele-assistance [1] to remote manipulation [2]. Mobile robots for teleconferencing are becoming increasingly commercialised [3]. Noteworthy examples are Anybots[1], TiLR[2] and Willow Garage's Texai[3]. However, most of these systems offer little more than a webcam on wheels, with the user being in control of every function of the robot. Our view is that a telepresence robot should be a tool for the user, not the main object of attention. This should hold for both experienced and novice users alike.

In this paper we investigate to what extent an increased level of robot autonomy will render the user's telepresence experience more pleasant. We focus on three major aspects of telepresence, namely i) robot control, ii) people interaction and iii) situation awareness. Robot control represents the initial problem which the user has to address. Most of the works in the literature provide a simple safety mechanism that slows down or stops the robot when it gets very close to an obstacle. This might render the user's experience very frustrating, as he/she might not be aware of the obstacles nearby the robot, given for example the difference between the robot sensors and the human view of the world. Our solution is to mediate the human commands with the robot perception of the surroundings, allowing the user to give only general directional commands while the robot's controller negotiates the obstacles in a smooth and danger-free way.

The user's view of the remote world is through the robot's camera. In our experiments the camera is attached to a pan-tilt unit (PTU) shaped like a robotic head (Figure 1). The main purpose of the PTU is to centre the view on the people with whom the user is interacting. This renders both participants' conversation more natural. Most of the time the user will not have to control the PTU. The robot is equipped with a face detection module that allows it to detect people in the surrounding environment and track their face if the user requests so. This way the user who is engaged in a conversation will have his/her interlocutor centred in the view all the times.

[1] http://anybots.com
[2] http://robodynamics.com/
[3] http://www.willowgarage.com/pages/robots/texai/overview

Situation awareness refers to the user's perception of the robot surroundings and status [4]. In our application we provide the user with a 3D-like representation of the environment around the robot. In particular the user interface shows a map of the environment (when available), the current and the past laser readings and the position of the people around the robot. This provides the user with an intuitive representation of the robot interface without being overwhelmed by details not concerning his/her scope, which is to interact with people.

We tested our proposed system with 8 users, all of whom had never previously used a mobile robot. We asked every user to complete a questionnaire about the experience with and without the additional autonomy levels. The results show that the user's experience is greatly enhanced if the robot manages several low-level tasks.

The main application of the proposed system is in teleconferencing. A visitor to an office could remotely connect to the robot and meet people engaging them in conversation. Other applications include homecare assistance and virtual tourism.

The main contribution of this paper is in the study of robot-assisted communication. From the user point of view the interlocutor is semi-automatically selected and always centred in the interface. From the interlocutor point of view the robot's head always keeps eye contact while talking. This renders the conversation and the overall experience pleasant for both parties. We present a user-centric testing of robot-assisted interaction in teleoperation [5].

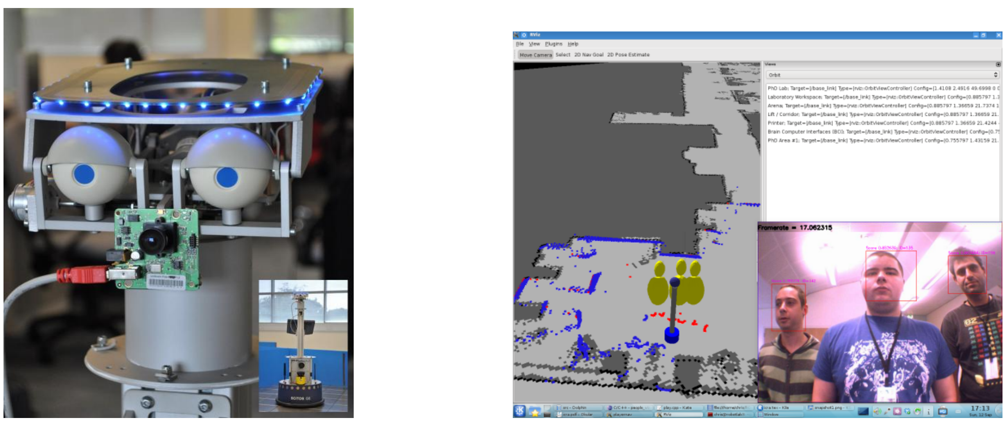

Fig. 1. (Left) The robot used in the experiments. A firewire camera with a 90° lens is attached to a pan-tilt head unit. This in turn is attached to a MetraLabs Scitos-G5 base, with a laser range scanner mounted $25cm$ from the floor. (Right)The graphical user interface. The left half of the interface presents the user with a 3D representation of the robot surroundings, the location of the robot in the environment and the location of people around the robot. Clicking a location on the map instruct the robot to autonomously drive to that location. The lower right corner of the interface presents the camera feed, and highlights detected people. Clicking a person instructs the robot to autonomously track the person with the pan-tilt unit, thus aiding conversation.

2 Related Work and Motivations

Robotic telepresence has been deeply studied in the context of health care [6], multi-robot teleconferencing [7] and industrial vehicle control [8] among others. One of the most recent and most comprehensive studies of teleoperation for homecare assistance is in [1]. The authors designed a robotic platform suitable for operation in home environments where carpets, stairs, furniture and changes in the floor surface are the main obstacles around which, or over which, a robot must navigate.

Most research in telepresence focused mainly on the control problem. The teleoperated robot is usually located far from the user, and communication happens through the available telecommunication infrastructure, which most of the time uses a wireless channel. This introduces delays in the control loop that render the user's experience frustrating.

In [9] a collaborative control is introduced. The user's commands are represented as a directional vector in a potential fields based controller. The weight of these commands changes as the efficiency of the overall movements changes. While this approach is appealing for teleoperation, it is mostly suitable when the user can directly see the robot surroundings, like in the wheelchair application the authors used as a testbed. Moreover the proposed controller does not take into consideration the robot kinematics, which could cause non-smooth movements of the platform.

In [10] a mediation agent observes the robot state and decides, using symbolic knowledge, if a teleoperated action is safe. The authors recognise that the proposed approach requires powerful reasoning capabilities and decided to test only obstacle avoidance and goal-seeking capabilities. Although the idea of reasoning about the user's commands instead of providing a reactive controller is appealing, we believe that the robot performance will be affected by this approach rending the teleoperation unresponsive and the whole user's experience unpleasant.

In [11] the authors provide several levels of autonomy, from full autonomy to complete teleoperation. Our proposed system is positioned below full autonomy, in that the user shares control with the robot. As noted by the authors, this approach yields the best robot effectiveness provided that the user does not neglect the robot. As we are concerned only with a single-robot application, our approach complies with this requirement. While the authors investigated only autonomy in robot movements, we expanded the study by considering semi-autonomy in interaction and situation awareness as well.

Semi-autonomous interaction with surrounding people has been much less studied. In [12] a robot that copies the user's height and head orientation is introduced. The authors also introduced a basic system to maintain eye contact and gaze. A different approach is in [13], where a museum tour guide robot capable of interaction with multiple users is presented. However, the robot perception and attentional control is controlled by a human operator, thus increasing the user's cognitive demand to run the robot.

The above systems are still more robot-centric than user-centric. The user has to constantly monitor obstacles close to the robot, as often the only aid in the navigation is the robot slowing down or stopping when too close to an obstacle (see [1] among others). This is not an easy process as the user relies mainly on vision to navigate, while a robot usually detects obstacles using a range sensor. Understanding the robot's sensors therefore becomes mandatory to successfully operate the robot.

A second barrier is represented by the robot's camera. A human constantly moves his/her head and eyes to keep contact with the interlocutor. This is a very important aspect of social interaction, especially during conversations. Reproducing the same movements with a robotic head is a demanding task, as the user has to control several degrees of freedom of the robot. Ignoring this aspect leads to unnatural interaction between the robot and the surrounding people [5].

These considerations led us to perform a study on what types of aids a robot can provide to the user in a teleoperation scenario. We therefore equipped the robot with the capability to detect people in the environment and to maintain eye contact during the conversation. Moreover the presence of people around the robot is clearly represented in the user interface, thus increasing the situation awareness. The robot is therefore semi-autonomous in control and interaction, thus allowing the user to primary focus only on the system's task: telepresence.

3 Techniques

3.1 Assisted Control

The robot's assisted control is based on the well-known dynamic window controller (DWC) [14]. According to this approach the robot's linear velocity v and angular velocity ω are chosen to maximise the objective function $G(v, \omega)$ defined in equation (1).

$$G(v, \omega) = \alpha \cdot angle(v, \omega) + \beta \cdot dist(v, \omega) + \gamma \cdot velocity(v, \omega) \tag{1}$$

where α, β, γ are three constant values, $angle()$ is a function that reaches the maximum when the angle between the robot and the desired goal is minimum, $dist()$ is a function which reaches the maximum when the robot chooses a collision-free course and $velocity()$ is equal to the chosen v. The search space for v, ω is bounded by the dynamical constraints of the robot motion (hence the term "dynamic window"). The controller is assumed to issue commands every dt seconds.

The user controls the robot using a joystick whose x and y axis are linearly mapped to the desired v and w. We substituted the $velocity()$ function with the desired v, and the $angle()$ function is substituted by $\omega \cdot dt$. In the original DWC approach the maximum of $G(v, \omega)$ is calculated by discretising the search space and querying the value of each cell. This might lead to unnecessary calculations and it could disrupt the real-time requirements of the DWC, as the effective number of queries depends on the particular processor's load. We therefore adopted a Monte Carlo approach to the optimisation of G, evaluating uniformly random v, ω pairs for dt seconds and storing the pair that yields the best v, ω value. On the robot's on-board PC this allows for around 1500 queries guaranteed to be carried out within the dt time limit and evenly spaced in the search space.

The constant values α, β and γ heavily influence the robot performance. Some of these values might lead to the robot not avoiding collision (for example an high α) or to an unstable motion. Their value depend on the dynamics of the platform being used and there is no general rule to set them. Extensive testing was performed with and without experienced users to find the best combination that allows the robot to follow the user's commands while staying safe. As a result we found the best combination being $\alpha = 0.3$, $\beta = 0.2$ and $\gamma = 0.5$.

The above modifications to the DWC mean that the robot is not given a goal in the environment to reach, but a series of rotational and translational velocities to keep. These velocities are combined with the robot sensing to obtain the final v and ω. As we will show in the next section the users learnt that most of the time steering the joystick is not necessary as the robot's path will smoothly adapt to the surrounding environment. This is turn fulfills the goal of reducing the cognitive workload necessary to control the robot resulting in a more pleasant experience with the teleoperation. Moreover this approach to control limits the effect of communication delays often seen in teleoperation.

3.2 People Detection and Tracking

The main novelty of our proposed system is the semi-autonomous robot-interlocutor interaction, which relies on the robot being able to detect people in the environment. The robot is equipped with the Viola-Jones face detection module [15]. Using this approach people can be detected up to $2m$ away from the robot. We tested the face detection module in several different environmental and light conditions, and the resulting confusion matrix is reported in Table 1. Although the detector is reliable, sometimes it

Table 1. Confusion matrix for the Viola-Jones classifier.

	Present	Not Present
Detected	95%	1.7%
Not detected	5%	98.3%

occasionally flickers between detection and non-detection, and its speed is very limited at 3 frames per second. To sidestep these problems every time a new face is detected a tracker is initialised. Details are as follows.

– Faces are detected in the camera image. The output of the face detector is an image region, I_f.
– Each time a face F is detected, a region tracker is initialised around the face region. If the detected face overlaps any existing tracked region, then the tracker associated with the existing region is terminated.
– Detected image regions are tracked within the image (in $2D$) at approximately 20 frames per second. The tracking is carried out by finding the best matching location for S (the matching score) in the current image within a region around the previous location. Matching is performed using normalised cross correlation of the grey level pixel values, according to equation 2.

$$S(x, y) = \frac{\sum_{x',y'} (F(x', y') \cdot I_f(x + x', y + y'))}{\sqrt{\sum_{x',y'} (F(x', y')^2 \cdot \sum_{x',y'} I_f(x + x', y + y')^2)}} \tag{2}$$

  where I_f is the search area in the original image, which is twice the size of the face template, F is the face template to track and S is the tracking score. The location with the highest score indicates the next position for the tracker, provided that the score is greater than a threshold $c = 0.75$. Otherwise the tracker is deleted.

The approximate 3D location of the detected people is then found by fusing the laser range finder with the camera data. To achieve this the backprojection vector of the pixel in the centre of the persons' face is transformed into the laser co-ordinate frame. The horizontal angle of this vector then provides the approximate bearing in the laser scan at which the legs of the person should be located. The closest distance in a $5°$ region around the bearing to the legs is used to provide the range to the person.

Although the 3D location is only approximate, it is accurate enough to provide the end user with a meaningful display of the locations of people. Furthermore, it allows some false positives (i.e non-face objects detected as faces) to be eliminated from the tracking by deleting trackers associated with people that are either too tall or too distant.

3.3 User Interface

Figure 1 shows the interface we developed for this application. It presents the user with both a $3D$ view of the environment and the $2D$ camera view. In the 3D view the robot is depicted along with the positions of people surrounding it. A map of the environment is given in grey, with the current laser scan superimposed on the map. This example demonstrates that the 3D display increases the user's situation awareness. When using only the camera feed the user's view is blocked by people, but by incorporating the map and the people locations the user is provided with an intuitive display of a wider area surrounding the robot. Furthermore, this helps the user to understand the laser scan readings shown in red and blue.

The interface allows the user to navigate the robot through the environment using either the joystick, or by selecting a desired location on the map. Clicking a face within the 2D image instructs the robot to keep that face centred within the image by moving the pan-tilt unit. Clicking outside of a face region returns the pan-tilt unit to the centre position. This means that the user is not required to manually control the pan-tilt head unit, thus reducing the total workload.

We extensively used open source software for this project. The user interface is built on top of the Robotic Open Source project[4]. Voice and video streaming are performed using Ekiga[5] libraries.

4 Experimental Results

The goal of the experiments was to assess the improvement of the users' teleoperation experience provided by the enhancements described above. We conducted tests to validate the three major aspects of our work, i.e.:

- The assisted control described in section 3.1, which we anticipated would require the user only to give the robot broad heading controls, while obstacle avoidance is carried by the remote robot low-level controller.
- The people detection and gaze control described in section 3.2 which was designed to permit the user and the interlocutors to have natural conversation and improve both sides' experience.
- The user interface described in section 3.3, intended to improve the situation awareness of the user and his/her ability to direct the robot where required.

While the first point can be numerically evaluated, the two other points are based on the user's subjective judgment. Therefore we evaluated the users' satisfaction using the well-established Witmer and Singer questionnaire on telepresence [16]. This questionnaire provides a well-established way to measure the performance of a remotely controlled agent in a virtual environment. We found this questionnaire to be the most suitable for our application, as all the questions are relevant to the task the user had to accomplish.

Of eight users, half used the robot with the autonomy enhancements while the other half used unassisted control without people detection or gaze control. None of them were told they were using a different system. Before every test the user was allowed 10 minutes of training to familiarise themselves with the system and the goals. At the end of both tests we asked the users to complete the questionnaire. During all the experiments the users had no visual contact with the robot, thus using only the interface.

4.1 Evaluation of Assisted Control

The goal of the first test was to measure the number of corrections the users had to provide to the robot via the joystick. To measure this we asked the users to follow a S-shaped path in a cluttered office-like environment. Four users were using the assisted control while the other four had full control over the robot. Figure 2 shows two example trajectories followed by the robot with and without the assisted control.

We defined a correction as any joystick lateral change larger than 5%, in accordance to [9]. Table 2 summarises the number of corrections performed with and without the assisted control. It can be seen that the assisted control requires significantly less corrections than plain control, thus reducing the user's involvement and stress.

[4] http://www.ros.org
[5] http://ekiga.org/

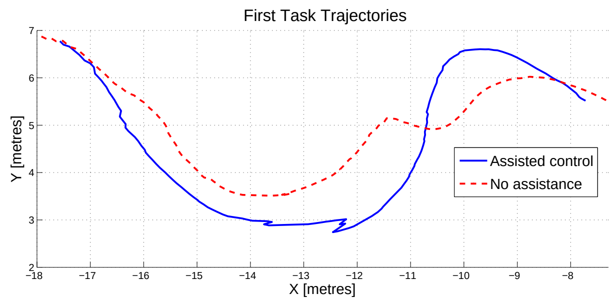

Fig. 2. Two example user trajectories, one with assisted control and one without. Note that some discontinuities in the plotted trajectories are the result of small localisation errors in the map-based navigation system.

Table 2. Number of corrections for each of the 8 users.

Assisted Control	No Assistance
33	211
27	298
41	288
38	196

4.2 Evaluation of the User Satisfaction

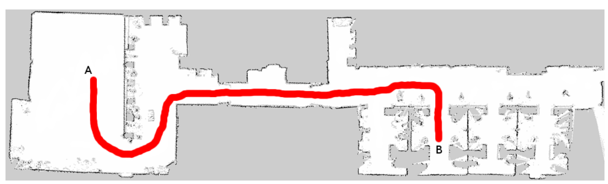

Fig. 3. The map of the environment where the experiments were conducted. The users had to drive the robot from point A to point B and back. The entire path length is around 60m. The robot footprint is represented by the width of the red path shape.

In the second test we asked the users to navigate the robot through a series of rooms, locate a group of people and engage in conversation. The route is represented in Figure 3. The robot started from point A and the interlocutors were located at point B. The robot had to traverse a long narrow corridor and drive among several desks. The environment was again a cluttered office-like space. Several desks, chairs and other items are located in the environment. At the end of every test the user was asked to fill a Witmer and Singer questionnaire. We deliberately performed one test for every user (apart from the initial training) to avoid any bias possibly introduced by experience. None of the users had previously controlled a mobile robot before. The users' ages ranged between 20 and 35 years old mixed male-female people.

The results in Figure 4 show a significant improvement of the users' satisfaction while using the enhanced autonomy.

5 Discussion

Here we discuss a few questions that strongly support our proposed approach.

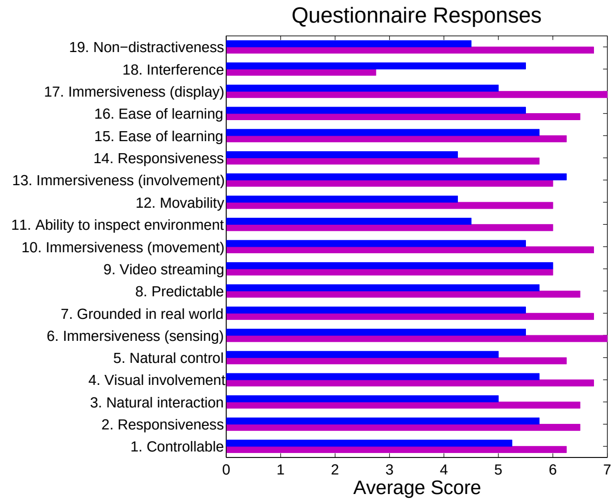

Fig. 4. Results of the user satisfaction questionnaire. For every row, the top blue line represents the average score for the users using no assistance, while the bottom magenta line represents the average score for the users using the enhanced robot autonomy. The text for the questions is a summary of the original question, detailed in Section 7

.

- The answers to the questions $1, 3$ and 5 shows that users are more comfortable with the robot-assisted control. Although this result is not new in literature, it is noteworthy since having a natural control over the robot was one of the main goals of this approach.
- The answers to question 10 show that the users felt more immersed when provided with a $3D$ representation of the robot's surroundings. In particular we believe that the higher score is due to the representation in the interface of the people around the robot.
- We were expecting the answers to question 14 to be equal, as the delay involved in the communication between the users' computer and the robot were the same with and without enhancements. We believe the the higher score is due to the assisted control mitigating the influence of delays.
- The answers to question 17 show that, although our interface presents the user with far more information than a plain camera view, it did not interfere or distract with the primary goal of this system, i.e. interacting with people.
- The answers to question 19 further support the above points.

In addition to the qualitative user satisfaction, the experiment described in section 4.1 numerically shows the improvements in the trajectory followed by the users while navigating in a S-shaped environment. In particular, Figure 2 highlights the differences between the manually and assisted control. In the case of manual control the user had to correct the trajectory several times, often to compensate for the effects of delayed control. The assisted control user was able to smoothly navigate in the environment. Table 2 shows the number of corrections each user had to perform to follow the trajectory in Figure 2. The 4 users that were using the assisted control needed a significantly less number of corrections compared to the users who were manually driving the robot.

6 Conclusions and Future Work

This paper presented a robotics system for teleoperation with semi-autonomous control and interaction. The user only exerts control over a few high level tasks, while the robot manages all the low level details. A contribution of this paper is in the assessment and verification that, in a teleoperation scenario, not only assistance in control is necessary, but also assistance in the interaction with the interlocutors. This has been achieved by providing the robot with people detection and tracking capabilities that allow it to maintain eye contact with the interlocutor. We showed in several experiments that the users' satisfaction is greatly enhanced by the the robot's semi-autonomy.

In an office or home environment obstacles exist at various heights, therefore being not always visible by the robot's laser range finder. In the future we plan to provide the robot with 3D sensing capabilities in order to safely navigate in these environments. Moreover during our tests we found that some users had difficulties locating a speaker when engaged in a conversation with several users. We believe this is due to the lack of the directional sound sensing a human being is capable of. We therefore plan to integrate a stereo microphone and incorporate sound localisation techniques so as to enhance the situational awareness and the overall user experience.

7 Appendix 1: The questionnaire

The questionnaire presented to the users asked 19 of the Witmer and Singer [16] virtual presence questionnaire questions, each receiving a score from 1 to 7. A score of 1 represents a negative response to the question, while a score of 7 represents a positive response. In all questions except 18, the higher the user response the more satisfied they are. The questions asked were as follows:

1. How much were you able to control events?
2. How responsive was the environment to actions that you initiated (or performed)?
3. How natural did your interactions with the environment seem?
4. How much did the visual aspects of the environment involve you?
5. How natural was the mechanism which controlled movement through the environment?
6. How compelling was your sense of objects moving through space?
7. How much did your experiences in the virtual environment seem consistent with your real world experiences?
8. Were you able to anticipate what would happen next in response to the actions that you performed?
9. How completely were you able to survey or search the environment with your vision?
10. How compelling was your sense of moving around inside the virtual environment?
11. How closely were you able to examine objects?
12. How well could you examine objects from multiple positions?
13. How involved were you in the virtual environment experience?
14. How much delay did you experience between your actions and the expected outcomes?
15. How quickly did you adjust to the virtual environment?
16. How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?
17. How much did the visual display quality interfere or distract you from performing assigned tasks or required activities?
18. How much did the control devices interfere with the performance of assigned tasks or with other activities?
19. How well could your concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?

8 Acknowledgements

The authors thank Dr. Yianni Gatsoulis for his help with the Witmer and Singer questionnaire and discussions. Our thanks are also to all the users that participated to the experiments and contributed with their comments. This research has been supported by the Centre of Excellence in Intelligent Systems (CoEIS) project, funded by Northern Ireland Integrated Development Fund and InvestNI.

References

1. F. Michaud, P. Boissy, D. Labonté, S. Brière, K. Perreault, H. Corriveau, a. Grant, M. Lauria, R. Cloutier, and M.-a. Roux, "Exploratory design and evaluation of a homecare teleassistive mobile robotic system," *Mechatronics*, pp. 1–16, Apr. 2010.
2. K. Shiratsuchi, K. Kawata, E. Vander Poorten, and Y. Yokokohji, "Design and Evaluation of a Telepresence Vision System for Manipulation Tasks," *Proceedings 2007 IEEE International Conference on Robotics and Automation*, no. April, pp. 4313–4318, Apr. 2007.
3. M. Ingebretsen, "Where's My Personal Robot?" *Intelligent Systems, IEEE*, vol. 24, no. 6, pp. 90–93, 2009.
4. H. Yanco and J. Drury, "Where am I? Acquiring situation awareness using a remote robot platform," in *IEEE Conference on Systems, Man and Cybernetics*, 2004, pp. 2835–2840.
5. K. Nakadai, K. Hidai, H. Mizoguchi, H. Okuno, and H. Kitano, "Real-Time Auditory and Visual Multiple-Object Tracking for Humanoids," in *Proceedings of the 17th international joint conference on Artificial intelligence-Volume 2*. Morgan Kaufmann Publishers Inc., 2001, pp. 1425–1432.
6. K. K. Chung, K. W. Grathwohl, R. K. Poropatich, S. E. Wolf, and J. B. Holcomb, "Robotic Telepresence: Past, Present, and Future," *Journal of Cardiothoracic and Vascular Anesthesia*, vol. 21, no. 4, p. 3, 2007.
7. N. Karnad, K. Ramakrishnan, and V. Isler, "A Multi-Robot System for Unconfined Video-Conferencing," *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pp. 355–361, 2010.
8. J. Larsson, M. Broxvall, and A. Saffiotti, "An Evaluation of Local Autonomy Applied to Teleoperated Vehicles in Underground Mines," *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pp. 1745–1752, 2010.
9. C. Urdiales, J. M. Peula, R. Annicchiaricco, F. Sandoval, and C. Caltagirone, "Efficiency based modulation for wheelchair driving collaborative control," *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pp. 199–204, 2010.
10. R. Wegner and J. Anderson, "Agent-Based Support for Balancing Teleoperation and Autonomy in Urban Search and Rescue," *International Journal of Robotics and Automation*, vol. 21, no. 2, pp. 1–19, 2006.
11. M. Goodrich, D. Olsen, J. Crandall, and T. Palmer, "Experiments in adjustable autonomy," *Proceedings of IJCAI Workshop on Autonomy, Delegation and Control: Interacting with Intelligent Agents*, 2001.
12. N. Jouppi and S. Thomas, "Telepresence Systems With Automatic Preservation of User Head Height, Local Rotation, and Remote Translation," *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, no. April, pp. 62–68, 2005.
13. M. Shiomi, T. Kanda, S. Koizumi, H. Ishiguro, and N. Hagita, "Group attention control for communication robots with wizard of OZ approach," *Proceeding of the ACM/IEEE international conference on Human-robot interaction - HRI '07*, p. 121, 2007.
14. D. Fox, W. Burgard, and S. Thrun, "The dynamic window approach to collision avoidance," *IEEE Robotics & Automation Magazine*, vol. 4, no. 1, pp. 23–33, 1997.
15. P. Viola and M. J. Jones, "Robust Real-Time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, May 2004.
16. B. Witmer and M. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Presence*, vol. 7, no. 3, pp. 225–240, 1998.